ДОПОВІДІ
НАЦІОНАЛЬНОЇ
АКАДЕМІЇ НАУК
УКРАЇНИ

7 • 2016

ІНФОРМАТИКА ТА КІБЕРНЕТИКА

http://dx.doi.org/10.15407/dopovidi2016.07.027

УДК 519.7

О. А. Галкін

Київський національний університет ім. Тараса Шевченка
E-mail: galkin.o.a@gmail.com

Дослідження властивостей функцій екстраполяційної глибини з використанням ядерних оцінок щільності

(Представлено членом-кореспондентом НАН України А. В. Анісімовим)

Розроблено математичний апарат, призначений для дослідження та розв'язання багатокласових задач класифікації на основі використання функцій екстраполяційної глибини. Досліджено властивості функцій екстраполяційної глибини, що дозволяють отримати стійкий до викидів непараметричний класифікатор, здатний обходити об'єкти з нульовою глибиною. Вивчаються непараметричні критерії для визначення та побудови багаторівневої структури згладжування, що дозволяє отримати глобальні властивості функцій щільності та меж класів при відповідних умовах регулярності.

Ключові слова: функція екстраполяційної глибини, байєсівський ризик, ядерні оцінки щільності.

Постановка задачі. Як відомо, класифікатор максимальної глибини класифікує досліджуваний об'єкт за класом, відносно якого він знаходиться на максимальній глибині. Враховуючи складність обчислень, використання довільних функцій глибини для класифікації даних на основі максимальної глибини може бути дещо ускладненим у просторі великої розмірності. Зокрема, функція глибини ∇° є легко обчислювальною, однак не є афінно-інваріантною. Для вирішення цієї проблеми необхідно використовувати оцінки з гарантованим рівнем значущості для матриці розсіювання Ξ_H, де афінно-інваріантність задається так:

$$\nabla^\circ E(z, H) = 1 - \left\| \Omega_H \left\{ \frac{\Xi_H^{-1/2}(z - Z)}{\|\Xi_H^{-1/2}(z - Z)\|} \right\} \right\|. \tag{1}$$

Класифікатор максимальної екстраполяційної глибини задається у вигляді

$$\Im_1(z) = \arg \max_{l \in \{1,\ldots,L\}} F_e(z, H_{lm_l}) = \arg \min_{l \in \{1,\ldots,L\}} \Upsilon(z, H_{lm_l}), \tag{2}$$

де $m_1, m_2, \ldots, m_L$ — об'єкт з L конкуруючих класів, а H_{lm_l} — емпірична функція розподілу L-го класу для $l = 1, \ldots, L$. Зазначимо, що $F_e(z, H_l)$ є монотонно зростаючою функцією від $h_l(z)$, якщо множинна функція щільності h_l є еліптично-симетричою та унімодальною [1].

Властивості функцій максимальної екстраполяційної глибини. Можна стверджувати, що множинна форма класифікатора $\Im_1(\cdot)$ збігається з байєсівським класифікатором $\Im_B(z) = \arg \max\limits_{1 \leqslant l \leqslant L} p_l h_l(z)$, якщо $p_1 = \ldots = p_L = 1/L$, тобто коли апріорні ймовірності конкуруючих класів є рівними, а також, якщо h_l відрізняються лише своїми параметрами розташування. Тому при збільшенні розміру вибірки даних частота помилок класифікатора $\Im_1(\cdot)$ також сходиться до байєсівського ризику Ψ_B, тобто частоти помилок $\Im_B(\cdot)$.

Лема 1. *Якщо для деякої функції щільності h та параметрів розташування ε_1, ε_2, $\ldots$, ε_L функції h_1, h_1, $\ldots$, h_L задовольняють модель зсуву розташування, тобто $h_l(z) = h(z - \varepsilon_l)$, а також є унімодальними та еліптично-симетричними, то частота помилок класифікатора максимальної екстраполяційної глибини $\Im_1(\cdot)$ сходиться до байєсівського ризику при* $\min\{m_1, m_2, \ldots, m_L\} \to \infty$.

Доведення. Як слідує з леми 1, множинна форма класифікатора $\Im_1(\cdot)$ є байєсівським класифікатором. Отже ми маємо

$$|\Psi(\Im_1) - \Psi_B| \leqslant$$
$$\leqslant \frac{1}{L} \sum_{l=1}^{L} \int \left| \prod_{i=1, i \neq l}^{L} \Lambda \left\{ \frac{F_e(z, H_{lm_l})}{F_e(z, H_{im_i})} > 1 \right\} - \prod_{i=1, i \neq l}^{L} \Lambda \left\{ \frac{F_e(z, H_l)}{F_e(z, H_i)} > 1 \right\} \right| h_l(z)\, dz, \qquad (3)$$

оскільки частота помилок $\Im_1(\cdot)$ задається, як $\Psi(\Im_1) = \sum\limits_{l=1}^{L} p_l P\{\Im_1(Z) \neq l\}$, де $Z \in l$ — класу. Таким чином, для $l = 1, \ldots, L$ має місце така збіжність:

$$\sup_{z \in \mathbb{R}^r} |F_e(z, H_{lm_l}) - F_e(z, H_l)| \overset{\text{ac.}}{\to} 0 \qquad \text{при} \qquad m_l \to \infty. \qquad (4)$$

В результаті, використовуючи теорему Лебега про мажоровану збіжність, можна стверджувати, що $\Psi(\Im_1) \to \Psi_B$. Лему доведено.

Теорема 1. *Нехай множинний розподіл H має еліптично-симетричну щільність з параметрами розташування та масштабу ε та Ξ відповідно. Якщо V_H є визначеною постійною величиною, тоді має місце така рівність:*

$$C(z, H) = \{(z - \varepsilon)' \Xi^{-1} (z - \varepsilon)\}^{1/2} = V_H \Upsilon(z, H).$$

Доведення. Враховуючи симетричність $\varphi' Z$ відносно $\varphi' \varepsilon$, ми маємо, що $\varepsilon_H(\varphi' Z) = \varphi' \varepsilon$ для $\forall \varphi \in \mathbb{R}^r$. Звідси

$$\Upsilon(z, H) = \sup_{\varphi\colon \|\varphi\|=1} \left\{ \frac{|\varphi' z - \varepsilon_H(\varphi' Z)|}{\eta_H(\varphi' Z)} \right\} = \sup_{\varphi\colon \|\varphi\|=1} \left\{ \frac{|\varphi'(z - \varepsilon)|}{\Sigma_H(\varphi' Z)} \frac{\Sigma_H(\varphi' Z)}{\eta_H(\varphi' Z)} \right\}, \qquad (5)$$

де Σ_H — стандартне відхилення.

Можна стверджувати, що $\varphi' X \overset{r}{=} \|\varphi\| X_1$, де X_1 — перша компонента $X = (X_1, X_2, \ldots, X_r)'$. Даний результат має місце, оскільки для $\forall \varphi \in \mathbb{R}^r$, $X = \Xi^{-1/2}(Z - \varepsilon)$ є сферично-розподіленою функцією. Таким чином, ми маємо $\varphi' Z = \varphi' \varepsilon + \varphi' \Xi^{1/2} Z = \varepsilon_\varphi + j'_\varphi X \overset{r}{=} \varepsilon_\varphi + \|j_\varphi\| X_1$, де $\varepsilon_\varphi = \varphi' \varepsilon$ та $j_\varphi = \Xi^{1/2} \varphi$.

Отже, $\eta_H(\varphi' Z) = \|j_\varphi\| \eta_H(X_1)$, а $\Sigma_H(\varphi' Z) = \|j_\varphi\| \Sigma_H(X_1)$. Звідси

$$\frac{\Sigma_H(\varphi' Z)}{\eta_H(\varphi' Z)} = \frac{\Sigma_H(X_1)}{\eta_H(X_1)} = \frac{1}{V_H}. \tag{6}$$

Отже, результат має місце, оскільки

$$\sup_{\varphi\colon \|\varphi=1\|} \left\{ \frac{|\varphi'(z-\varepsilon)|}{\Sigma_H(\varphi' Z)} \right\} = \{(z-\varepsilon)' \Xi^{-1} (z-\varepsilon)\}^{1/2}, \tag{7}$$

а також враховуючи той факт, що V_H не залежить від φ. Теорему доведено.

У даній роботі досліджуються властивості функцій екстраполяційної глибини, що головним чином базуються на еліптичній симетрії розподілів. Розглянемо випадок, коли розподіли множин даних є еліптичними. Якщо $E(z, H_l)$ є глибиною z відносно H_l, тоді байєсівський класифікатор задається так:

$$\Im_B(z) = \arg \max_{1 \leqslant l \leqslant L} p_l o_l \{E(z, H_l)\}, \tag{8}$$

де o_l — відповідна функція перетворення. Зауважимо, що функція перетворення o_l є монотонно-спадною та однаковою для всіх груп множин даних, якщо функції h_l є унімодальними, а розподіли множин даних відрізняються лише параметрами розташування відповідно [2]. Крім того, байєсівський класифікатор є еквівалентним класифікатору максимальної глибини, якщо p_l рівні. Проте, при невиконанні хоча б одного з вищенаведених припущень виникає потреба в отриманні інформації щодо функціональних форм o_l. Для функцій екстраполяційної глибини функції перетворення o_l можуть бути отримані на основі теореми 1.

Лема 2. *Якщо $\xi_l(\cdot)$ є функцією щільності від $F_e(z, H_l)$, а функції h_1, h_2, $\ldots$, h_L є еліптично-симетричними, то байєсівський класифікатор задається, як*

$$\Im_B(z) = \arg \max_{l \in \{1,\ldots,L\}} \frac{\alpha_l \xi_l \{F_e(z, H_l)\} \{F_e(z, H_l)\}^{r-3}}{\{1 - F_e(z, H_l)\}^{r-1}}, \tag{9}$$

де α_l — постійна величина.

Доведення. Враховуючи еліптичну симетрію функцій h_l, ми маємо

$$h_l(z) = \mathrm{I}\left(\frac{r}{2}\right) (2p)^{-r/2} |\Xi_l|^{-1/2} \frac{c_l(C(z, H_l))}{C(z, H_l)^{r-1}}, \tag{10}$$

де c_l — ймовірнісна функція щільності від $C(z, H_l) = \{(z-\varepsilon_l)' \Xi_l^{-1} (z-\varepsilon_l)\}^{1/2}$, а ε_l та Ξ_l — параметри розташування та масштабу для h_l відповідно.

Отже, можна стверджувати, що

$$\Im_{\mathrm{B}}(z) = \arg \max_{1 \leqslant l \leqslant L} p_l h_l(z) = \arg \max_{1 \leqslant l \leqslant L} \frac{\alpha_l \lambda_l \{\mathrm{I}(z, H_l)\}}{\{\mathrm{I}(z, H_l)\}^{r-1}}, \tag{11}$$

де постійна величина α_l залежить від H_l та p_l, а λ_l є функцією щільності від $\mathrm{I}(z, H_l)$. Зазначимо, що класифікатор (11) слідує з теореми 1. Враховуючи той факт, що

$$F_e(z, H_l) = \{1 + \mathrm{I}(z, H_l)\}^{-1}, \tag{12}$$

доведення випливає з властивостей вибіркового розподілу. Лему доведено.

Як слідує з доведення теореми 1, поки α_l змінюються залежно від вибору одновимірних мір розташування та масштабу, лема 2 є справедливою для довільного визначення функцій екстраполяційної глибини.

Ядерні оцінки щільності. Для побудови розвиненої версії класифікатора екстраполяційної глибини ми використовуємо метод ядерних оцінок щільності для оцінювання ξ_l та вибіркову форму $F_e(z, H_{lm_l})$ для оцінювання $F_e(z, H_l)$. У даному випадку оцінюється лише одновимірна щільність незалежно від розмірності простору вимірювань. Таке оцінювання дозволяє обійти проблему "прокляття розмірності", що часто має місце в багатовимірних непараметричних оцінках щільності.

Зауважимо, що вибір смуги пропускання a_l є обов'язковим для оцінки ξ_l, де $1 \leqslant l \leqslant L$. Дана оцінка щільності задається так:

$$\overline{\xi}_{la_l}(\omega) = (m_l a_l)^{-1} \sum_{i=1}^{m_l} \Theta\{a_l^{-1}(\omega - \overline{\omega}_{m_l}^{(l)}(z_{li}))\}, \tag{13}$$

де Θ — функція ядра, а $\overline{\omega}_{m_l}^{(l)}(z) = F_e(z, H_{lm_l})$.

У даному випадку визначений рівень згладжування може демонструвати різну поведінку у різних областях простору вимірювань. Виходячи з цього, актуальною задачею є дослідження результатів класифікації для різних масштабів згладжування замість використання фіксованої пари (a_1, a_2) у визначеному діапазоні [3]. Об'єднання даних, що індексовані по смугах пропускання, можна проводити за допомогою прийняття зваженого середнього значення оцінених апостеріорних ймовірностей [4].

Зазначимо, шо $e^{\rho_{m,a_1,a_2}(z)}$ дає оцінку $p_1 h_1(z)/p_2 h_2(z)$, оскільки елемент z класифікується до першого класу, якщо

$$\rho_{m,a_1,a_2}(z) = \log[d_{m_1,a_1}^{(1)}(z)] - \log[d_{m_2,a_2}^{(2)}(z)] - \Delta > 0, \tag{14}$$

де Δ вибирається шляхом мінімізації помилки перехресної перевірки для фіксованих (a_1, a_2). Отже, ми маємо

$$\overline{\pi}_{m,a_1,a_2}(1|z) = \frac{e^{\rho_{m,a_1,a_2}(z)}}{1 + e^{\rho_{m,a_1,a_2}(z)}}, \tag{15}$$

що є оціненою апостеріорною ймовірністю класу.

Оскільки $\pi_m^*(l|z) = \sum\limits_{a_1,a_2 \in A} q_{a_1,a_2} \overline{\pi}_{m,a_1,a_2}(l|z)$, результуючий класифікатор отримується шляхом об'єднання апостеріорних оцінок, отриманих при різних значеннях (a_1, a_2), а саме:

$$\Im_3(z) = \arg\max_{l=1,2} \pi_m^*(l|z). \tag{16}$$

Зазначимо, що q_{a_1,a_2} — вага, що присвоюється класифікатору, для якого a_1 та a_2 є смугами пропускання двох класів [5].

Теорема 2. *Припустимо, що для* $l = 1, 2$, h_1 *та* h_2 *є еліптично-симетричними, де* $h_l(z) > 0$ *для* $\forall z \in \mathbb{R}^r$, *а* $H_{\beta,l}(\bar{z}) = P(\beta(Z) \leqslant \bar{z})$ *є рівномірно неперервною функцією в* $\bar{z}$, *де* $\beta(z) = d^{(2)}(z)/d^{(1)}(z)$, $d^{(l)}(z) = \omega_l(\xi^{(l)}(z))(\xi^{(l)}(z))^{r-3}/(1 - \xi^{(l)}(z))^{r-1}$, $\xi^{(l)}(z) = F_{np}(z, H_l)$, *а* $Z \in l$*-му класу. Також припустимо, що для* a_l^g *та* a_l^j *мають місце збіжності* $a_l \to 0$ *та* $m_l a_l^4 \to \infty$ *при* $m_l \to \infty$. *Тоді коефіцієнт помилкової класифікації багаторівневого класифікатора екстраполяційної глибини* $\Im_3(\cdot)$ *сходиться до байєсівського ризику при* $\min\{m_1, m_2\} \to \infty$.

Доведення. Доведення даної теореми слідує з теореми Лебега про мажоровану збіжність при умові, якщо для фіксованого z, $\pi^*_m(1|z) \xrightarrow{P} \pi(1|z)$ при $\min\{m_1, m_2\} \to \infty$.

Припустимо, що збіжність $\pi^*_m(1|z) \xrightarrow{P} \pi(1|z)$ не має місця. Отже, $\exists\{m_\Delta = (m_{1\Delta}, m_{2\Delta})\colon \Delta \geqslant 1\}$ та $\mu_0 > 0$, що для $\forall\Delta \geqslant 1$, $|\pi^*_{m_\Delta}(1|z) - \pi(1|z)| > \mu_0$. Нехай $\{A_{m_\Delta}\}$ — відповідна послідовність діапазону смуг пропускання, де $\Delta \geqslant 1$.

Враховуючи той факт, що $\pi^*_{m_\Delta}(1|z)$ є зваженим середнім значенням $\bar{\pi}_{m_\Delta, a_1, a_2}(1|z)$, можна отримати таку підпослідовність $\{(a_1^{m_\Delta}, a_2^{m_\Delta}) \in A_{m_\Delta}, \Delta \geqslant 1\}$, що $|\bar{\pi}_{m_\Delta, a_1^{m_\Delta}, a_2^{m_\Delta}}(1|z) - \pi(1|z)| > \mu_0$ для всіх $\Delta \geqslant 1$. Звідси слідує, що збіжність $\bar{\pi}_{m_\Delta, a_1^{m_\Delta}, a_2^{m_\Delta}}(1|z) \xrightarrow{P} \pi(1|z)$ не має місця. В результаті маємо протиріччя, оскільки послідовність смуг пропускання задовольняє умову регулярності, при якій для $l = 1, 2$, $a_l \to 0$ та $m_l a_l^4 \to \infty$ при $m_l \to \infty$. Теорему доведено.

Підводячи підсумки можна стверджувати, що вибір вагової функції q не має значного впливу на вибіркову продуктивність класифікатора $\Im_3(\cdot)$. Однак вибір A та q є необхідним при використанні скінченної вибірки. В результаті, необхідно використовувати більш великі ваги для класифікаторів, що мають більш низькі частоти помилок, де вага повинна поступово зменшуватися при збільшенні частоти помилок.

Цитована література

1. *Rousseeum P. J., Struyf A.* Characterizing angular symmetry and regression symmetry // Statist. Plann. Inference. – 2004. – **122**. – P. 163–170.
2. *Pollard D.* Convergence of Stochastic Processes. – New York: Springer, 1984. – P. 1–10.
3. *Галкин А. А.* Применение мер сглаживания в непараметрических ядерных классификаторах с использованием нормальной аппроксимации вероятностей // Проблемы управления и информатики. – 2015. – № 5. – С. 85–92.
4. *Zuo Y. J.* Projection-based depth functions and associated medians // Annals of Statistics. – 2003. – **31**. – P. 1463–1484.
5. *Chacon J. I., Duong T., Wand M. P.* Asymptotics for general multivariate kernel density derivative estimators // Statist. – 2011. – **21**. – P. 810–837.

References

1. *Rousseeum P. J., Struyf A.* Statist. Plann. Inference, 2004, **122**: 163–170.
2. *Pollard D.* Convergence of Stochastic Processes, New York: Springer, 1984.
3. *Galkin O. A. Problems of Control and* Informatics, 2015, No 5: 85–92 (in Russian).
4. *Zuo Y. J.* Annals of Statistics, 2003, **31**: 1463–1484.
5. *Chacon J. I., Duong T., Wand M. P.* Statist, 2011, **21**: 810–837.

Надійшло до редакції 26.10.2015

А. А. Галкин

Киевский национальный университет им. Тараса Шевченко
E-mail: galkin.o.a@gmail.com

Исследование свойств функций экстраполяционной глубины с использованием ядерных оценок плотности

Разработан математический аппарат, предназначенный для исследования и решения многоклассовых задач классификации на основе использования функций экстраполяционной глубины. Исследованы свойства функций экстраполяционной глубины, позволяющие получить *устойчивый к выбросам непараметрический классификатор, способный обходить объекты с нулевой глубиной. Исследованы непараметрические критерии для определения и построения многоуровневой структуры сглаживания, которая позволяет получить глобальные свойства функций плотности и границ классов при соответствующих условиях регулярности.*

Ключевые слова: функция экстраполяционной глубины, байесовский риск, ядерные оценки плотности.

O. A. Galkin

Taras Shevchenko National University of Kiev

E-mail: galkin.o.a@gmail.com

Research of properties of the extrapolation depth functions using kernel density estimates

The mathematical apparatus, which is intended for the research and solving the multiclass classification problems based on the use of the extrapolation depth functions, is developed. The properties of the extrapolation depth functions, which allow one to obtain some nonparametric classifier resistant to outliers that is able to bypass objects with zero depth, are studied. Nonparametric criteria to define and to build a multilevel smoothing structure, which enables one to obtain the global properties of density functions and the boundaries of classes under appropriate conditions of regularity, are studied.

Keywords: extrapolation depth function, bayesian risk, kernel density estimates.